Geophys. J. Int. (2005) **160,** 179–194 doi: 10.1111/j.1365-246X.2005.02460.x

Temporal clustering of major earthquakes along individual faults due to post-seismic reloading

Shelley J. Kenner* and Mark Simons

Seismological Laboratory, MC 252-21, *California Institute of Technology, Pasadena, CA* 91125, *USA. E-mail: simons@caltech.edu*

Accepted 2004 July 26. Received 2004 April 26; in original form 2003 November 17

SUMMARY

Palaeoseismic evidence suggests that earthquake recurrence intervals in some regions can be highly variable, with clusters of multiple large events separated by much longer periods of quiescence. Because post-seismic processes have a significant effect on the reloading rate of the coseismic fault, we hypothesize that temporal variations in the amount of stress concentrated in the non-seismogenic lithosphere can modulate large earthquake recurrence times. We explore this hypothesis using simple analogue spring-dashpot-slider models. We find that in the presence of small amounts of environmental noise, post-seismic stress transfer over timescales much longer than an earthquake cycle may be an important factor in generating clustering behaviour. The propensity for the system to be clustered is a function of a non-dimensional number that we call the Wallace Number, W. W is defined as the average earthquake stress drop divided by the product of the long-term geologic strain rate across the fault and the effective viscosity of the system. Our results indicate that environments with relatively low strain rates and a relatively weak non-seismogenic lithosphere are most susceptible to clustering driven by post-seismic stress recycling mechanisms.

Key words: earthquakes, fault models, post-seismic, rheology, viscoelasticity.

1 INTRODUCTION

Analysis of worldwide earthquake catalogues indicates that large earthquakes may cluster in time (Kagan & Knopoff 1976; Kagan & Jackson 1991). Many of these clustered earthquakes are spatially distributed within a particular fault system (e.g. Wallace 1987; King *et al.* 1994; Goes 1996; McCalpin & Nishenko 1996; Xu & Deng 1996; Stein *et al.* 1997; Rockwell *et al.* 2000). Palaeoseismic evidence demonstrates that major earthquakes may also form temporal clusters along a single fault segment. Individual faults with highly irregular recurrence intervals include the Dead Sea transform fault and faults in the Basin and Range province, among others (e.g. Wallace 1987; Swan 1988; Grant & Sieh 1994; Ritz *et al.* 1995; Marco *et al.* 1996; Xu & Deng 1996; Dorsey *et al.* 1997; Friedrich *et al.* 2003). In particular, palaeoseismological observations along the Wasatch fault, on the eastern boundary of the Basin and Range, over a number of different timescales and intervals suggests that major earthquakes occur in clusters with intracluster repeat times of a few thousand years separated by intercluster periods of tens of thousands of years (Wallace 1987; Friedrich *et al.* 2003).

For the purposes of this study, we somewhat arbitrarily define major earthquake clustering along a single fault as the occurrence of multiple event sequences with intracluster inter-event times that are many times smaller than the average time between clusters. Many processes may contribute to clustering of this type, including along strike rheological heterogeneities, fault zone constitutive behaviour, fault segmentation, fault interactions and fault healing rate. For certain prescribed timescales of fault healing, models incorporating damage rheology predict mode switching between periods of high seismicity with few large earthquakes and periods of low seismicity with characteristic earthquakes (Ben-Zion *et al.* 1999; Lyakhovsky *et al.* 2001). Over multiple event timescales, we will show that, in noisy environments, time-dependent post-seismic processes may also play an important role in the generation of major earthquake clusters.

Over shorter timescales, it is apparent from observed surface velocity transients following great earthquakes that time-dependent post-seismic processes can represent a measurable component of the deformation field for many decades after the coseismic event. Examples include the 1906 M_w 7.9 San Francisco earthquake (Thatcher 1983; Kenner & Segall 2000), the 1944 and 1946 Nankai Trough earthquakes (Thatcher 1984) and the 1964 M_w 9.2 Alaska earthquake (Li & Kisslinger 1985; Savage & Plafker 1991). Post-seismic stress transfer typically reinforces the coseismic perturbation. Post-seismic stress transfer is especially significant at the ends of ruptures, potentially leading to temporally clustered cascades of earthquakes along the length of a fault. In an elastic context, cascade-type behaviour has been proposed by Stein *et al.* (1997) for the North Anatolian fault with any temporal delay explained using laboratory

*Now at: Department of Geological Sciences, 101 Slone Building, University of Kentucky, Lexington, KY 40506-0053, USA. E-mail: skenner@uky.edu

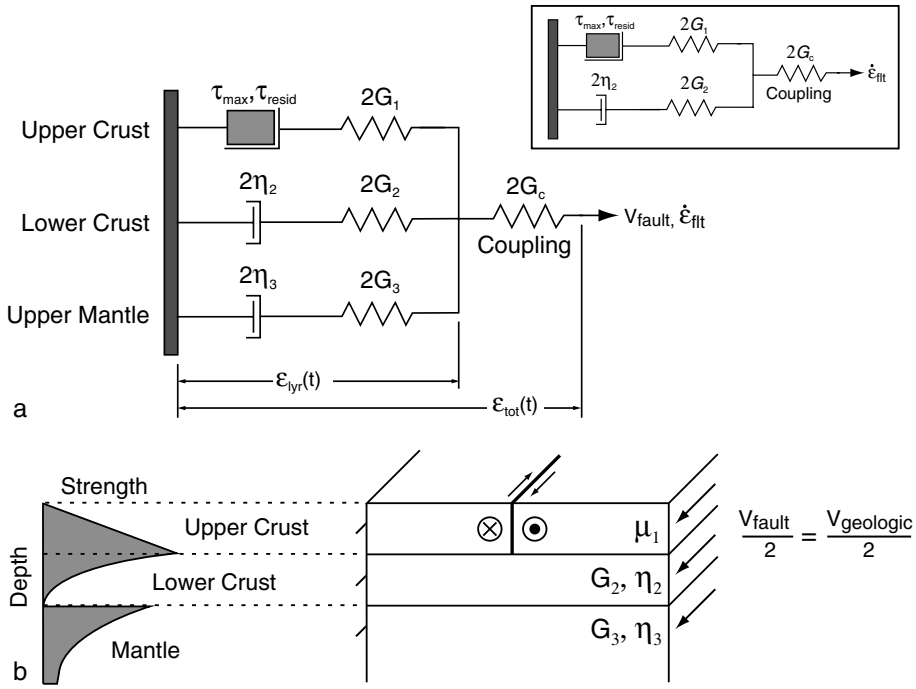

Figure 1. (a) A spring-dashpot-slider model designed to approximate the mechanical interactions that take place in a three-layer Earth-like system with (b) a faulted elastic plate overlying a viscoelastic channel above a viscoelastic half-space. Insert in (a) shows the equivalent two-layer model.

based rate-and-state friction laws (e.g. Dieterich 1994; Harris & Simpson 1998). Spatially clustered sequences of major earthquakes on tectonically related faults, including the Landers, Big Bear, Joshua Tree sequence modelled elastically by King *et al.* (1994), the Eastern California shear zone (Rockwell *et al.* 2000), the Indian–Asian collisional zone (e.g. Xu & Deng 1996; Chéry *et al.* 2001), and the Basin and Range province (e.g. Wallace 1987; McCalpin & Nishenko 1996), may have a similar explanation. The possibility that post-seismic processes can generate temporal clusters of spatially distributed major earthquakes has been demonstrated analytically using the two-fault spring-dashpot-slider system of Chéry *et al.* (2001) for timescales comparable to a single earthquake repeat time. Multiplet earthquakes along adjoining subduction zone segments are modelled by Nomanbhoy & Ruff (1996) using a two-asperity spring-frictional slider-creeping slider model over multiple event time intervals. Lynch *et al.* (2003) use continuum finite element models to show that viscoelastic coupling can produce clustered earthquakes in a simple two-fault system, however none of these examples considers post-seismic reloading of the original coseismic fault.

Post-seismic phenomena reload the coseismic fault at a rate that can initially be much faster than the background rate predicted from far-field plate motions (e.g. Lehner & Li 1982; Thatcher 1983; Rundle 1986; Li & Rice 1987; Kenner & Segall 1999). As the non-seismogenic lithosphere relaxes earthquake induced stress concentrations, a portion of the load is shed upwards to the seismogenic crust and coseismic fault. As a result, part of the stress transferred to the non-seismogenic lithosphere during the coseismic event is effectively recycled. Post-seismic stresses may therefore comprise a substantial portion of the total stress required to initiate failure in the next major earthquake (Kenner 2004). To date, the influence of post-seismic stress recycling on the coseismic fault has only been addressed at timescales comparable to a single earthquake repeat time.

To investigate the influence of post-seismic phenomena on the recurrence rate of major earthquakes on a single fault segment over longer timescales, we present an analogue spring-dashpot-slider model of time-dependent mechanical interactions analogous to those that may take place in the Earth (Fig. 1). With this tool, we gain a greater understanding of how stress is transferred throughout the lithospheric system with time and how variations in rheology and loading rate affect the system behaviour. In particular, we investigate how the system responds to environmental noise, where sources of noise could include natural variations in the fault zone rheology, fault zone heterogeneity and stress transients resulting from earthquakes on neighbouring faults. With no noise, the system generates regularly repeating earthquakes. Conversely, in noisy conditions we show that the fundamental behaviour of the analogue fault system can change significantly.

2 MODEL OVERVIEW

Many authors have developed spring-slider and other types of cellular automaton models to investigate earthquake statistics and observed earthquake frequency-magnitude distributions (e.g. Burridge & Knopoff 1967; Rundle 1988; Carlson & Langer 1989; Turcotte *et al.* 2000). Some studies have also included viscous dashpots in an effort to account for time-dependent effects (e.g. Hainzl *et al.* 1999, 2000; Pelletier 2000). In each case, sliders represent a single fault subpatch and are combined in large systems via springs and/or dashpots. Depending on the forces exerted, individual patches are allowed to fail independently or in simultaneous cascades of many patches. Because our goal is to consider recurrence intervals between large earthquakes having approximately the same magnitude, we make no attempt in this study to generate Gutenburg–Richter frequency–magnitude statistics or foreshock/aftershock sequences.

Our spring-dashpot-slider system (Fig. 1a) simulates, by analogy, the stress interactions that take place in a three-layer model

composed of an elastic plate overlying a viscoelastic channel above a viscoelastic half-space. The elastic layer contains a single antiplane, vertical strike-slip fault (Fig. 1b). The entire system is loaded via a constant applied velocity, $V_{\rm flt}$, which represents the long-term average geologic slip-rate of the fault. Layer 1 represents the faulted elastic crust. Deformation in the non-seismogenic lithosphere, represented by layers 2 and 3, is approximated using time-dependent Maxwell rheologies. Layer 2 represents the lower crust above the Moho. Layer 3 represents the upper mantle. Where noted, a simpler two-layer model (Fig.1a inset) is also considered. Note, however, that the spring-dashpot-slider system is a 1-D system. Thus, direct comparisons to specific tectonic environments are unwarranted.

The coupling spring, denoted by the subscript c, regulates interseismic stress transfer between the three layers of the analogue model, effectively transforming the kinematic boundary condition, $V_{\rm flt}$, to a stress boundary condition that transiently reacts to post-seismic stress perturbations. As the stiffness of the coupling spring becomes infinite, the model layers behave independently. There is no post-seismic stress transfer between layers and each layer strains at a constant rate, $V_{\rm flt}$. As the stiffness of the coupling spring approaches zero, all applied strains are accommodated by the coupling spring and no stress is transferred to layers 1–3. In essence, the coupling spring increases the dimension of the 1-D spring-dashpot model. In 2-D finite element models, a coupling spring would not be needed. Instead, the various rheological layers would be naturally coupled as part of a continuous medium.

As our objective is to explore interactions within simple layered viscoelastic systems and investigate variations in major earthquake recurrence intervals and clustering of main shocks on a single fault, the slider represents a single seismogenic fault with dimensions comparable to a characteristic earthquake in the tectonic region of interest. We are not attempting to model temporally clustered earthquakes that are spatially distributed, although many of the same processes may be acting. The slider is locked between earthquakes such that it behaves as a rigid element. The system is then loaded at a constant rate, $V_{\rm flt}$, until stresses in layer 1 reach some prescribed level. At this point the block slips, releasing accumulated stresses in layer 1. Because stress is conserved coseismically, stresses equivalent to the earthquake stress drop are transferred to the underlying viscoelastic layers during the slip event. Stress conservation during earthquakes also serves to increase the dimensionality of the system. Environmental noise is included in the system via small random variations in the fault (i.e. slider) failure criteria.

3 MODEL DERIVATION

Using the nomenclature in Table 1 and Fig. 1 and the governing equations for stress, τ, and strain, ε, in a spring, $\tau = 2G\varepsilon$, and dashpot, $\tau = 2\eta\dot{\varepsilon}$, the interseismic system behaviour can be described by the differential equations

$$\text{two-layer:}\quad \gamma_1\dot{\varepsilon}_{\rm lyr} + \gamma_0\varepsilon_{\rm lyr} = \delta_1\dot{\varepsilon}_{\rm tot} + \delta_0\varepsilon_{\rm tot}, \tag{1a}$$

$$\text{three-layer:}\quad \beta_2\ddot{\varepsilon}_{\rm lyr} + \beta_1\dot{\varepsilon}_{\rm lyr} + \beta_0\varepsilon_{\rm lyr} = \alpha_2\ddot{\varepsilon}_{\rm tot} + \alpha_1\dot{\varepsilon}_{\rm tot} + \alpha_0\varepsilon_{\rm tot}, \tag{1b}$$

where $\dot{\varepsilon}$ and $\ddot{\varepsilon}$ are the first and second time derivatives of strain, and β_i, α_i, δ_i, and γ_i are constant coefficients determined solely by the rheological parameters (Appendix A). For simplicity, only the two-layer equations will be given in the text. The analogous three-layer equations can be found in Appendix C. In fact, we will show that the governing equations for the three-layer model (Fig. 1a) can be recast as an equivalent two-layer model (Fig. 1a inset) given an appropriate choice of $\eta_{\rm eff}$.

Table 1. Nomenclature.

Variable	Definition
t	Time since the last earthquake
$G_{\rm eff}$	Effective lithospheric shear moduli
G_j	Shear modulus of spring j, where $j = 1, 2, 3$
$\eta_{\rm eff}$	Effective lithospheric viscosity
η_j	Viscosity of dashpot j, where $j = 1, 2, 3$
$T^m_{\rm eff}$	Effective lithospheric Maxwell relaxation time ($\eta_{\rm eff}/G_{\rm eff}$)
T^m_j	Maxwell relaxation time (η_j/G_j) of layer j, where $j = 1, 2, 3$
T^m_W	Effective system relaxation time as defined by W (4c)
$\varepsilon_{Dj}(t)$	Strain in dashpot j, where $j = 2, 3$
$\varepsilon_{Sj}(t)$	Strain across spring j, where $j = 1, 2, 3$
$\varepsilon_{\rm lyr}(t)$	Strain across any given layer
$\varepsilon_{\rm tot}(t)$	Strain in the coupling spring + $\varepsilon_{\rm lyr}$
$V_{\rm flt}$	Long-term fault slip-rate
$\dot{\varepsilon}_{\rm flt}$	Applied long-term strain accumulation rate of the system
$\tau_{\rm fail}$	Stress in element 1 when slip is initiated ($\tau_{\rm fail} \geq \tau_{\rm max}$)
$\tau_{\rm max}$	Nominal mean earthquake failure stress
$\tau_{\rm residual}$	Nominal mean residual fault stress following rupture
$N_{\rm max}$	Earthquake failure stress noise level (as a per cent of $\tau_{\rm max}$)
$N_{\rm residual}$	Residual fault stress noise level (as a per cent of $\tau_{\rm residual}$)
$\Delta\varepsilon_{\rm eq}$	Average earthquake strain drop
$\Delta u_{\rm eq}$	Average earthquake slip
$\Delta\tau_{\rm eq}$	Average earthquake stress drop
$T^{\rm eq}_{\rm avg}$	Average earthquake repeat time
$L_{\rm c}$	Characteristic distance for fault strain accumulation
W	Wallace number

Equivalent non-dimensional values are denoted with an asterisk.

For a constant strain-rate boundary condition, $\dot{\varepsilon}_{\rm tot} = \dot{\varepsilon}_{\rm flt}$, eq. (1a) reduces to

$$\gamma_1\dot{\varepsilon}_{\rm lyr} + \gamma_0\varepsilon_{\rm lyr} = (\delta_0 t + \delta_1)\dot{\varepsilon}_{\rm flt} + \delta_0\varepsilon_{\rm tot}|_{t=0}\,, \tag{2}$$

where the constant velocity boundary condition, $V_{\rm flt}$, has been reformulated in terms of strain rate to maintain dimensional consistency. The resultant long-term average strain rate is defined as $\dot{\varepsilon}_{\rm flt} = V_{\rm flt}/L_{\rm c}$, where $L_{\rm c}$ is the characteristic length scale for strain-accumulation across the fault. Eq. (2) can be solved analytically for $\varepsilon_{\rm lyr}$.

We non-dimensionalize eq. (2) using $\Delta\tau_{\rm eq}$, the average coseismic stress drop, $\eta_{\rm eff}$, the effective viscosity of the non-seismogenic lithosphere and $\dot{\varepsilon}_{\rm flt}$, the far-field tectonic boundary condition. Using this scheme, non-dimensional time, stress, stressing rate, strain and strain rate are defined as

$$\begin{aligned}
t^* &= \left(\frac{\Delta\tau_{\rm eq}}{\eta_{\rm eff}}\right)t,\\
\tau^* &= \frac{\tau}{\Delta\tau_{\rm eq}},\\
\dot{\tau}^* &= \frac{\eta_{\rm eff}}{(\Delta\tau_{\rm eq})^2}\dot{\tau},\\
\varepsilon^* &= \frac{\Delta\tau_{\rm eq}}{\eta_{\rm eff}\dot{\varepsilon}_{\rm flt}}\varepsilon,\\
\dot{\varepsilon}^* &= \frac{\dot{\varepsilon}}{\dot{\varepsilon}_{\rm flt}},
\end{aligned} \tag{3}$$

where asterisks denote non-dimensional variables. $\Delta\tau_{\rm eq}$ characterizes the coseismic behaviour of the fault and $\eta_{\rm eff}$ characterizes the

rate of viscous dissipation of post-seismic stress concentrations. Because there is only one fault in the model, $\dot{\varepsilon}_{\rm flt}$ equals the long-term average strain rate across the fault. As fault slip becomes continuous, $\Delta\tau_{\rm eq}$ approaches 0.

We combine $\Delta\tau_{\rm eq}$, $\dot{\varepsilon}_{\rm flt}$ and $\eta_{\rm eff}$ to define a non-dimensional parameter W, which we call the Wallace number after the work after Robert E. Wallace who initially proposed temporal grouping of slip events separated by quiescent periods in the Basin and Range province (Wallace 1987). W is defined as

$$W = \frac{\Delta\tau_{\rm eq}}{\dot{\varepsilon}_{\rm flt}\eta_{\rm eff}}. \tag{4a}$$

Knowing that $\Delta\tau_{\rm eq} = 2G_1\Delta\varepsilon_{\rm eq}$ and the average earthquake recurrence interval $T^{\rm eq}_{\rm avg} = \Delta\varepsilon_{\rm eq}/\dot{\varepsilon}_{\rm flt}$, W can also be written as

$$W = \frac{2G_1\Delta\varepsilon_{\rm eq}}{\dot{\varepsilon}_{\rm flt}\eta_{\rm eff}}, \tag{4b}$$

or equivalently

$$W = \frac{T^{\rm eq}_{\rm avg}}{T^{m}_{W}}, \tag{4c}$$

where $\Delta\varepsilon_{\rm eq}$ is the average strain released during an earthquake, $T^{\rm eq}_{\rm avg}$ is the average earthquake recurrence interval and $T^{\rm m}_{\rm W} = \eta_{\rm eff}/2G_1$ is the characteristic relaxation time of the system. When using eq. (4a), all times and strains in the governing equations are automatically normalized by $T^{\rm eq*}_{\rm avg}$ and $\Delta\varepsilon^{*}_{\rm eq}$, respectively (Appendix A). Hereafter, asterisks will represent non-dimensionalized and, where appropriate, normalized values.

The non-dimensional, normalized solution to eq. (2) has the form

$$\varepsilon^{*}_{\rm lyr} = \Pi_1[1 - \exp(-t^{*}/\Gamma)] + \Pi_2\exp(-t^{*}/\Gamma) + \Pi_3 t^{*}, \tag{5}$$

where Π_i and Γ are non-dimensional functions of the material parameters, W, $\varepsilon^{*}_{\rm tot}$ (0), and $\varepsilon^{*}_{\rm lyr}$ (0). Details of the derivation can be found in Appendix B. Based on eq. (5), the model behaviour is governed solely by (i) W, which incorporates the boundary conditions and (ii) various non-dimensional ratios of rheological parameters. Equations describing the stress and strain within each element of the two-layer model at any time are also given in Appendix B.

During the interseismic calculation, the fault failure criterion is monitored at fixed intervals, dt^{*}, of between 0.01 and 5, with the exact value depending on the range of earthquake recurrence intervals generated by a given set of parameters. When the stress in layer 1 reaches the slider failure stress,$\tau^{*}_{\rm max}$, coseismic slip of the slider is initiated. Because the time step is finite, the actual stress at failure is $\tau^{*}_{\rm fail}$, where $\tau^{*}_{\rm fail} \geq \tau^{*}_{\rm max}$. Once initiated, slip continues until stress in layer 1 drops to $\tau^{*}_{\rm residual}$, the residual stress after failure. To investigate the response of the system in the presence of small amounts of noise, $\tau^{*}_{\rm max}$ and $\tau^{*}_{\rm residual}$ values are normally distributed. The standard deviation of the noise distribution, $N_{\rm max}$ and $N_{\rm residual}$, is specified as a percentage of $\tau^{*}_{\rm max}$ and $\tau^{*}_{\rm residual}$ (see Appendix D for more information). Stress released via slip in the crust is transferred to the underlying viscoelastic elements, where it is added to the stress already supported by these layers (Appendix B). The slip event is assumed to occur instantaneously and the initial conditions of the system are reset as described in Appendix B.

Given the preceding derivation, the complete coseismic and interseismic evolution of the two-layer system can be parametrized by specifying W, $G_{\rm c}/G_1$, G_2/G_1, $\tau^{*}_{\rm max,avg}$, $\tau^{*}_{\rm residual,avg}$, $N_{\rm max}$ and $N_{\rm residual}$. For the three-layer system (Appendix C), G_3/G_1 and η_3/η_2 are also required. It should be noted that the relationship between the $\eta_{\rm eff}$ in the definition of W and model parameters η_2 and η_3 is determined by the problem geometry. For the two-layer model, $\eta_{\rm eff} = \eta_2$. Because the strains in layers 2–3 of the three-layer model are the same and stress is transferred to each layer instantaneously during the earthquake, $\eta_{\rm eff} = (\eta_2 + \eta_3)/2$, the arithmetic mean of the layer viscosities, in the three-layer case. In the real earth, where stresses must pass through each layer in succession, the precise relationship between $\eta_{\rm eff}$, η_2, and η_3 may be difficult to determine *a priori*. To limit the parameter space, complete earthquake stress drops are assumed such that $\tau^{*}_{\rm residual,avg}$ and $N_{\rm residual}$ are fixed with zero magnitude and $\tau^{*}_{\rm max,avg} = 1$. Unless otherwise noted, the ratios G_2/G_1 and G_3/G_1 are set to 1.

To maintain geophysical relevance, we require that:

(i) the long-term average strain rate of the slider must match the applied strain-rate boundary condition, $\dot{\varepsilon}_{\rm flt}$;

(ii) if no additional earthquakes occur, stresses in the elastic crust (layer 1) must increase monotonically with time and crustal strain rates must approach a steady-state value greater than zero and less than the applied geologic rate (assuming finite coupling spring stiffness); and

(iii) if no additional earthquakes occur, stresses in the non-seismogenic lithosphere (layers 2–3) must return to their steady-state value, as required for a Maxwell viscoelastic material with a constant strain-rate boundary condition.

By satisfying these criteria, the analogue model simulates mechanical interactions that are known to occur in 2-D, antiplane numerical models of a faulted elastic plate overlying a viscoelastic layered space. Because the physics are similar, we can use the spring-dashpot model to more fully understand post-seismic relaxation processes and to more efficiently explore the relevant parameter space. This approach also allows for the generation of major earthquake histories that include thousands of events upon which useful statistical tests can be employed. Because of the 1-D nature of the model, however, our results should be compared with real tectonic regimes in only very general terms.

Finally, it is useful to note that in the limit as $W \to 0$, the slider slips continuously and there are no discrete earthquakes. As a result, post-seismic transients disappear, thereby eliminating the ability of imposed noise to generate temporally clustered earthquake behaviour. In this instance, normally distributed input fault failure values produce normally distributed earthquake repeat times. Conversely, in the limit as $W \to \infty$, underlying viscoelastic layers can support no stress between earthquakes (see eqs B4 and C7 in the Appendix). As a result, post-seismic relaxation is instantaneous. Because the boundary displacement, $\varepsilon_{\rm tot}$, can not change instantaneously, some of the released post-seismic stress is recycled to the elastic crust, instantaneously moving the fault closer to failure. The remainder of the stress in the underlying viscoelastic layers is dissipated via instantaneous relaxation of the dashpots. For the $W \to \infty$ case, the remainder of the stress required for failure in the next earthquake must accumulate via the far-field boundary condition. During this stage, the system behaves like two springs in series (spring 1 and the coupling spring) driven with a constant strain-rate boundary condition.

4 THE COUPLING SPRING

Following an earthquake, stress within the perturbed system evolves with time as various layers approach their steady-state strain rates. Stresses transferred to layers 2–3 during the earthquake are relaxed as the Maxwell elements approach a constant stress. The excess stress is transferred to other lithospheric layers or dissipated via

viscous deformation of the dashpots. The stiffness of the coupling spring controls the amount of interlayer stress transfer.

As mentioned previously, if the stiffness of the coupling spring is infinite, the three layers are completely isolated from one another. There is no coupling between the elastic crust and underlying time-dependent materials. All post-seismic stresses are confined within the layer to which they were initially transferred during the earthquake and are dissipated viscously. Time-dependent variations in stressing rate within the seismogenic layer are eliminated and deformation rates are equal to the applied geological rate.

As the stiffness of the coupling spring decreases, coupling between the layers increases and the steady-state deformation rate in the seismogenic crust decreases below the applied geological rate, because a greater percentage of the applied strain-rate boundary condition is accommodated via extension of the coupling spring. To understand the role of the coupling spring in greater detail, consider the real Earth. Given the typical arc-tangent depiction of interseismic strain accumulation across a fault, note that, between earthquakes, points in the seismogenic crust at or near the fault lag behind points in the far field. This represents stored strain energy. If there is no coupling between crust and mantle (i.e. infinite G_c), strain energy is stored entirely in the crust (spring 1) and it is all released instantaneously during the earthquake. Alternatively, if the non-seismogenic lithosphere and elastic crust are coupled, points in the transition zone at the top of the non-seismogenic lithosphere also lag behind the long-term average. Thus, strain energy is also stored in the transition zone, represented, in this model, by the coupling spring. In consequence, steady-state deformation rates in the seismogenic crust (i.e. layer 1) are less than the long-term geologic average (i.e. the applied boundary condition). Because coseismic faulting does not extend into the transition zone and stress is conserved during the earthquake, accumulated strain energy in the coupling spring/transition zone is not released instantaneously by slip along the coseismic fault. Instead, it is released transiently over a period of time immediately following the event (Fig. 2d). Assuming long-term motion at the assigned geologic slip-rate, this post-seismic transient compensates for the interseismic crustal strain deficit created by the coupling spring drag effect.

Post-seismically, a decrease in the coupling spring stiffness means that a greater proportion of the coseismic stress drop can be transferred between layers. As a result, post-seismic effects become more significant in defining the overall behaviour of the model. When the Maxwell relaxation time $T_2^m < T_3^m$, post-seismic stress is initially transferred from layer 2 to layers 1 and 3 (Figs 2a–c). Eventually, layer 3 relaxes as well and all post-seismic stresses that have not been dissipated viscously are recycled to the seismogenic layer. Transient post-seismic stress transfer to layer 1 (Fig. 2a) and transient reductions in the strain energy stored in the coupling spring (Fig. 2d) produce crustal deformation rates that are initially higher than the long-term geological average (Fig. 2e). This compensates for steady-state deformation rates that are slower than average.

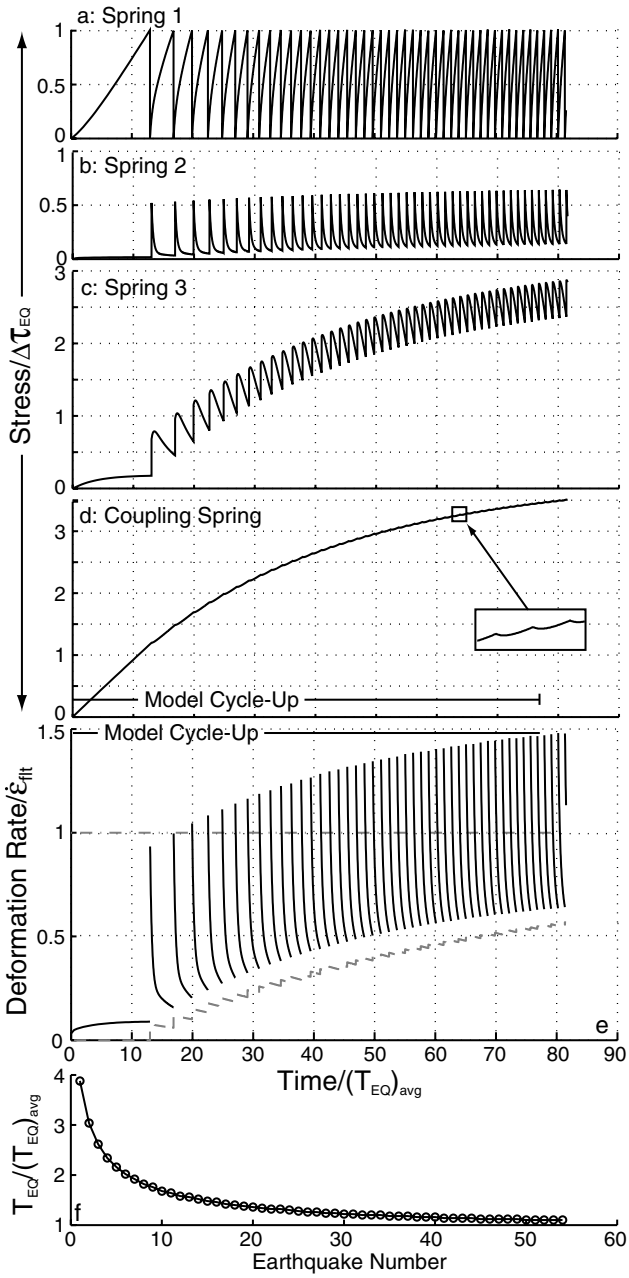

Figure 2. The temporal evolution of (a–d) stress in each element of the spring-dashpot model, (e) crustal strain rate (solid) and (f) earthquake recurrence intervals are shown as the system evolves to a steady-state condition. Material parameters are $W = 2$, $\eta_3/\eta_2 = 10$, $G_c/G_1 = 0.1$, $G_2/G_1 = 1$ and $G_3/G_1 = 1$ with no added noise ($N_{max} = 0$). At $t/T_{eq}^{avg} = 80$, the model cycle-up stage is nearing completion. Stress, recurrence intervals, etc. are nearing steady-state values. In (e), as time increases, the average fault strain release rate (dashed) asymptotically approaches the long-term geological rate (dash-dotted). Note that in (a–d), the stress scales are not uniform. The same parameter values, with noise, are used in Figs 7(a)–(c) and 8(a)–(f).

5 STRESS EVOLUTION DURING THE EARTHQUAKE CYCLE: NO NOISE

We first seek to understand the stress history in each of the layers within the spring-dashpot-slider system with no noise applied to the fault failure criteria (Fig. 2). Irrespective of rheological parameters, after an initial cycle-up period, the system reaches a steady-state condition with constant earthquake repeat times. The cycle-up period consists of two components. First, crustal strain rates and stresses in each non-seismogenic layer reach steady-state values representative of a state without earthquakes. Upon initiation of repeating earthquakes, if the recurrence interval is shorter than the total time required to relax post-seismic stresses in layers 2–3 (i.e. crustal strain rates do not return to their minimum, background value before the next event), the system must move to a second steady-state condition. Because peak strain accumulation rates in the crust are proportional to the magnitude of the post-seismic stress available

for transfer in the non-seismogenic layers (2–3), this new steady state represents a balance between the driving strain rate, $\dot{\varepsilon}_{\mathrm{flt}}$, the earthquake repeat time, the coseismic stress drop and the amount of post-seismic stress available in the non-seismogenic layers for recycling to the brittle crust. From an energy conservation perspective, steady-state conditions exist when the work done by tectonic loading is equal to the energy being dissipated by the viscous dashpots, where the rate of energy dissipation is proportional to the stress in the non-seismogenic layers. Peak crustal strain rates increase (Fig. 2e) and earthquake recurrence intervals decrease (Fig. 2f) until the new equilibrium state is reached. As exemplified by Lyzenga *et al.* (1991), Reches *et al.* (1994), Hager *et al.* (1999) and Lynch & Richards (2001), the presence of this cycle-up period demonstrates the importance of establishing a plausible background stress statall non-kinematic numerical models of earthquake faulting, especially if earthquake repeat times are less than the time required for complete post-seismic relaxation.

6 THE EFFECT OF NOISE

When noise is added to the maximum stress required for fault failure (Fig. 3), the resultant change in earthquake recurrence intervals is dependent on W and G_{c}. The skewness of the output recurrence interval distribution can be measured via the coefficient of variation, C_{v}, defined as the sample standard deviation/sample mean. For a purely periodic system, $C_{\mathrm{v}} = 0$. For a Poissonian process, $C_{\mathrm{v}} = 1$. Because the governing equations have been non-dimensionalized, the input τ^{*}_{max} has a $C_{\mathrm{v}} = \sim N_{\mathrm{max}}$ (Fig. 3). If clustering is not present, output recurrence intervals are also distributed with $C_{\mathrm{v}} = \sim N_{\mathrm{max}}$. Rigorously speaking, clustering is defined as $C_{\mathrm{v}} > 1$ (Kagan & Jackson 1991; Ben-Zion & Rice 1995), while the system is defined as quasi-periodic for $0 < C_{\mathrm{v}} < 1$. Clustering has also been defined on a more subjective basis that allows for qualitatively clustered behaviour with $C_{\mathrm{v}} < 1$ (McCalpin & Nishenko 1996). Our results indicate that earthquake repeat times become distinctly non-normal and temporal earthquake clustering becomes the dominate system behaviour when the maximum recurrence interval is ∼10 times larger than the typical intracluster recurrence interval. Although the evolution from normal to clustered is a continuous one, based on observations the transition to qualitatively clustered behaviour occurs

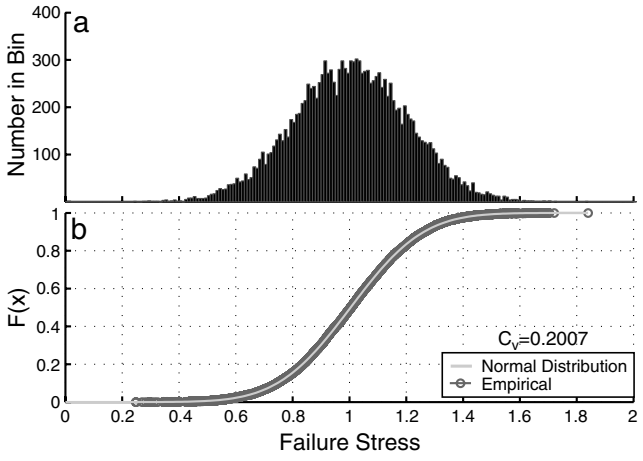

Figure 3. (a) Histogram and (b) empirical cumulative distribution function (CDF) for a representative distribution of τ^{*}_{max} values with normally distributed noise and $N_{\mathrm{max}} = 20$ per cent. This represents the input model noise for models shown in Figs 7 and 8. The histogram is comprises 0.01-MPa bins. The optimal normal CDF is superposed on the empirical CDF.

at an output C_{v} of roughly 0.5. For high viscosities (i.e. low W), prolonged model cycle-up periods can contaminate these statistical measures. We have removed earthquakes during the cycle-up period in order to obtain uncontaminated clustering statistics.

Whether or not the output recurrence intervals are qualitatively clustered, as W increases the system undergoes a distinct transition in state (Figs 4–6). Increasing W implies increasing stress drop, decreasing effective lithospheric viscosity and/or decreasing far-field tectonic strain accumulation rates. When W is small, the C_{v} of the output recurrence intervals is approximately equal to the C_{v} of the input noise. As W increases, the output C_{v} remains constant until some threshold W value is reached at the initiation of the transition zone. For all parameter combinations, as W increases beyond the transition zone, the output C_{v} again approaches a nearly constant value in what can be called the clustering regime. Assuming coupling between the seismogenic crust and non-seismogenic layers sufficient to allow some post-seismic stress transfer, the system behaviour is always within the clustered regime for large W. Note that, although the system state is clearly different and C_{v} has clearly increased, for certain parameter combinations, the output C_{v} in the clustering regime is <0.5 and does not produce qualitatively clustered behaviour (Figs 4–6).

With regard to other model parameters, we find that as the noise level increases, the amount of clustering increases (Fig. 4). Though the noise level affects the final C_{v} value in the clustered regime, it does not alter the W values bounding the transition zone (Fig. 4). The non-dimensionalized governing equations (5 and C3) suggest that changes in η_3/η_2 and depth variable shear moduli should have no effect on clustering behaviour. Fig. 5 shows that, if other parameters are held fixed, changes in η_3/η_2 in the three-layer model indeed have a no effect on clustering behaviour and all results collapse to a single curve. Despite small differences in the transition zone, Fig. 6 demonstrates that for changes in G_2/G_1 in the two-layer model, depth variable shear moduli also have a negligible effect on the clustering behaviour of the system.

To induce qualitatively clustered behaviour, we find that the ratio G_{c}/G_1 must be $\leq\sim 1$ (Fig. 6; a similar plot for the three-layer model would be identical). If the ratio is larger, the coupling spring stiffness limits transient post-seismic changes in $\varepsilon_{\mathrm{lyr}}$ and, as a result, post-seismic stress transfer to the seismogenic crust. When $G_{\mathrm{c}}/G_1 > 1$, post-seismic transients are not important and normally distributed noise on the failure criteria leads to nearly normally distributed recurrence intervals. Because G_2/G_1 and $G_3/G_1 \geq 1$, if $G_{\mathrm{c}}/G_1 < 1$, the most compliant element in the system becomes G_{c} and spring 1 now limits the post-seismic rate of change of $\varepsilon_{\mathrm{lyr}}$. As a result, post-seismic stress recycling to the seismogenic crust begins to be significant and output recurrence intervals begin to deviate from a normal distribution. For W in the clustering regime, as G_{c}/G_1 decreases, C_{v} values increase and short recurrence intervals dominate the statistics (Figs 6 , 7d–f and 8g–l). Furthermore, peak post-seismic crustal strain rates increase by ∼32 per cent as G_{c}/G_1 decreases from 1 to 0.1. In the non-clustered regime, G_{c}/G_1 ratios have a negligible effect on peak post-seismic strain rates.

We have found no effective methods for estimating realistic G_{c}/G_1 values, although $G_{\mathrm{c}}/G_1 < 1$ seems reasonable given observations of long-term post-seismic transients following major earthquakes (e.g. Thatcher 1984; Li & Kisslinger 1985; Savage & Plafker 1991). Observations following the 1906 $M = 7.8$ San Francisco earthquake, for example, give strain rates that are 2–3 times the geologic rate when averaged over the first two decades following the earthquake (Thatcher 1983; Kenner & Segall 2000). In 2-D or 3-D

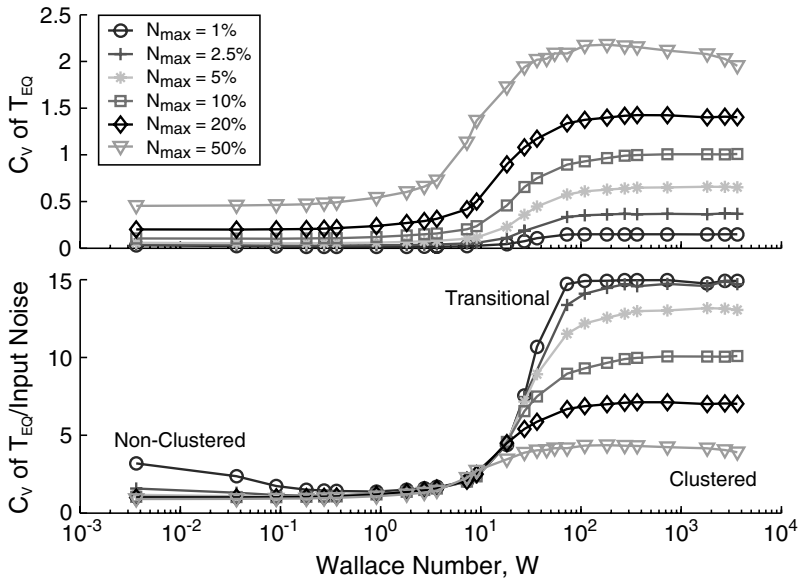

Figure 4. For the three-layer model, a semi-log plot of (a) the C_v of the output repeat times and (b) the C_v of the output repeat times normalized by the input C_v versus W as a function of noise level on the fault failure criterion, N_{max}. In all plotted models, $\eta_3/\eta_2 = 10$, $G_c/G_1 = 0.1$, $G_2/G_1 = 1$ and $G_3/G_1 = 1$. Note that, although the lowest N_{max} models are not qualitatively clustered for high W, the output C_v levels are much larger than the higher N_{max} models when normalized by the input C_v (compare a and b). The upward curvature as W decreases is the result of the prolonged cycle-up period necessary for models with very low W. Although earthquake recurrence intervals are not changing significantly over a few tens of earthquakes when these statistics were calculated, they are still getting progressively shorter over longer time periods. This biases the output C_v calculation. It was not numerically feasible to totally complete the cycle-up period for the low-W models prior to calculating their clustering statistics. This demonstrates the importance of understanding the cycle-up process in all multicycle numerical fault models. The effect is not as pronounced at higher noise levels because the increased system noise masks the continuing evolution of the average earthquake recurrence interval.

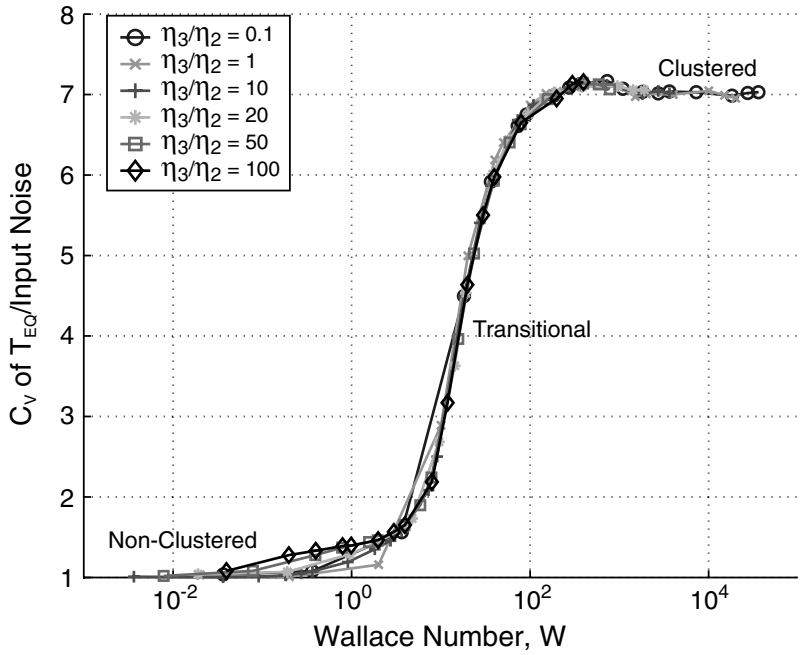

Figure 5. For the three-layer model, a semi-log plot of the C_v of the output repeat times normalized by the input C_v versus W as a function of ratio η_3/η_2, the upper mantle to lower crustal viscosity ratio. In all plotted models, $N_{max} = 20$ per cent, $G_c/G_1 = 0.1$, $G_2/G_1 = 1$ and $G_3/G_1 = 1$. Based on the fact that all of the curves overlap, it is clear that η_3/η_2 has no effect on the clustering behaviour of the system.

continuum models of the mechanical processes described here, a G_c value would not be required.

In this model, noise on the fault failure criterion approximates variations in the fault failure stress as a result of variations in fault healing and/or damage. Failure criteria noise also serves as a proxy for the temporal evolution of spatial heterogeneities and stress perturbations resulting from fault interactions. With this approach, clock advances and delays resulting from earthquakes on neighbouring faults are mapped into variations in the fault failure criterion. If fault interactions were explicitly included, applied noise

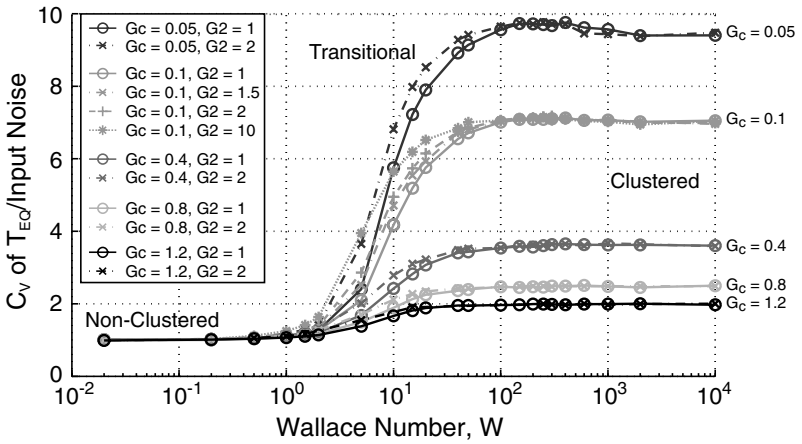

Figure 6. For the two-layer model, a semi-log plot of the C_v of the output repeat times normalized by the input C_v versus W for variations in the elastic moduli G_2 and G_c. The noise level on the fault failure criterion, N_{max}, is 20 per cent. Curves with the same G_c value plot nearly atop one another irrespective of the G_2 value. This is especially true within the clustered and non-clustered regimes. This implies that W is the primary parameter in determining whether or not earthquakes are clustered.

levels (20 per cent in Figs 7 and 8) could be lower and stress perturbations resulting from neighbouring earthquakes would provide the remaining variation necessary to generate clustering. If, in addition to the applied noise, fault interactions were explicitly included, clustering would be enhanced. Irrespective of the noise source, lowering the noise level to 5 per cent with $G_c/G_1 = 0.1$ and $\eta_3/\eta_2 = 10$ still produces clustered behaviour ($C_v = \sim 0.64$) for W in the clustering regime (Fig. 4).

7 IMPLICATIONS IN VARIOUS HYPOTHETICAL TECTONIC ENVIRONMENTS

At this point, we make some rough estimates of W in different, hypothetical tectonic environments. Because the analogue model is inherently 1-D, direct comparisons to specific faults systems are not warranted. This model should be viewed as an intuition building tool and not a predictive tool. Thus, direct comparisons will be left to future 2-D and 3-D continuum finite element models. Only general trends and characteristics will be discussed.

Using $\dot{\varepsilon}_{flt} = V_{flt}/L_c$, we assume L_c is equal to twice the seismogenic depth. V_{flt}, the long-term geologic fault slip rate, can be obtained from palaeoseismology. For the purposes of this rough calculation, assume a seismogenic depth of 15 km, $\Delta\tau_{eq}^{hypo} = 10$ MPa, $\eta_3^{hypo}/\eta_2^{hypo} = 10$ and a depth independent shear moduli, G_{eff}^{hypo}, of 6.4×10^4 MPa. Along the San Andreas fault, the effective Maxwell relaxation time, T_{eff}^{m}, is ~30 yr (Thatcher 1983; Li & Rice 1987; Kenner & Segall 2000). Assume this is characteristic of typical high strain-rate plate boundary regions. This yields an $\eta_{eff}^{hypo} = T_{eff}^{m} G_{eff}^{hypo}$ of $\sim 6 \times 10^{19}$ Pa s. For $V_{flt} = 35$ mm yr^{-1} (also characteristic of the San Andreas fault), this yields $W^{hypo} = \sim 4.5$. For $W^{hypo} = 2$, $G_c^{hypo} = 0.1$ and $N_{max}^{hypo} = 20$ per cent (Figs 7a–c and 8a–f), output recurrence intervals are distributed with $C_v = \sim 0.27$. As suggested by the fact that $W^{hypo} = 2$ places the system on the non-clustered side of the transition zone (Figs 4–6), visual inspection of the system behaviour suggests that temporal clustering of earthquakes is not a dominant process (Figs 8a–f). For this set of hypothetical parameters, we therefore conclude that temporal clustering of large earthquakes as a result of post-seismic stress recycling is minimal.

In contrast, if η_{eff}^{hypo}, G_c^{hypo} and N_{max}^{hypo} are not changed and V_{flt} is dropped to 3 mm yr^{-1}, $W^{hypo} = \sim 50$. Using $W^{hypo} = 40$, which places the model along the clustered side of the transition zone (Fig. 4), an output $C_v = \sim 1.173$ is obtained (Figs 7d–f and 8g–l). Visual inspection of model stress and strain histories (Figs 8g–l) indicates that temporal clustering dominates the overall behaviour of the system. $V_{flt} = 3$ mm yr^{-1} is a reasonable value for slowly deforming zones such as the Dead Sea transform (Klinger *et al.* 2000, and references therein) and the Basin and Range province (Niemi *et al.* 2004). If these regions also possess a weak non-seismogenic lithosphere, which is likely (Lowry & Smith 1995; Al-Zoubi & ten Brink 2002), η_{eff}^{hypo} would be smaller and W^{hypo} would be larger making clustering even more likely. For reference, palaeoseismic estimates of C_v for the Dead Sea transform fault are between 0.99 and 1.53 (Marco *et al.* 1996).

Temporal clustering of earthquakes as a result of post-seismic stress recycling may also have implications for earthquake recurrence in intraplate seismic zones. For instance, 1811–1812 type earthquakes in the New Madrid seismic zone in the south-central USA have occurred every ~ 500 yr (Tuttle *et al.* 2002) but evidence suggests that this cluster of faulting is limited to the Holocene (Pratt 1994; Schweig & Ellis 1994; Van Arsdale 2000). Though intraplate viscosities are 2–3 orders of magnitude higher than the η_{eff}^{hypo} used above, earthquake stress drops in intraplate regions are typically higher and intraplate strain rates are significantly lower, leading to W^{hypo} values in the clustered regime. For comparison, strain rates along the San Andreas fault and in the Basin and Range are $\sim 10^{-6}$–10^{-7} yr^{-1} (Thatcher 1990) and $\sim 10^{-8}$ yr^{-1} (Niemi *et al.* 2004), respectively. Long-term average seismic strain rates in the central and eastern USA are estimated to be between $\sim 10^{-11}$–10^{-12} yr^{-1} (Anderson 1986).

8 THE EARTHQUAKE STRESS BUDGET

As demonstrated, the interaction between long-term fault slip rates and non-seismogenic layer viscosities is very important in determining the overall behaviour of the system. This interaction reflects the role of recycled post-seismic stresses in reloading the coseismic fault. Neglecting system noise, the expected recurrence interval is inversely proportional to the applied long-term strain rate across the fault (eqs. 4abc). As the time required for complete relaxation of the non-seismogenic layers becomes long relative to the expected recurrence interval (i.e. low-W environments with high viscosity

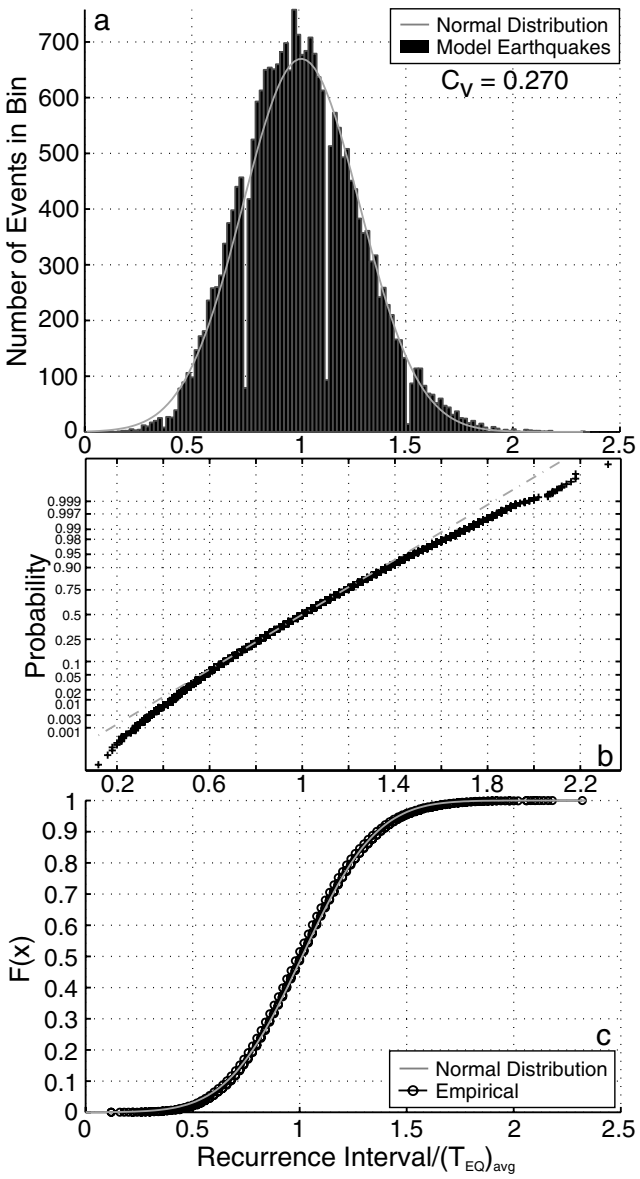

Figure 7. (a, d) Histogram, (b, e) normal probability plot and (c, f) empirical cumulative distribution function (CDF) of earthquake recurrence intervals for models with (a–c) $W = 2$ and (d–f) $= 40$. In both cases, $\eta_3/\eta_2 = 10$, $G_c/G_1 = 0.1$, $G_2/G_1 = 1$, $G_3/G_1 = 1$ and $N_{max} = 20$ per cent. Histograms are composed of $dt^*/T^{eq^*}_{avg}$ bins. In (b, e), if output recurrence intervals are normally distributed, the normal probability plot (crosses) will lie along the straight dashed line.

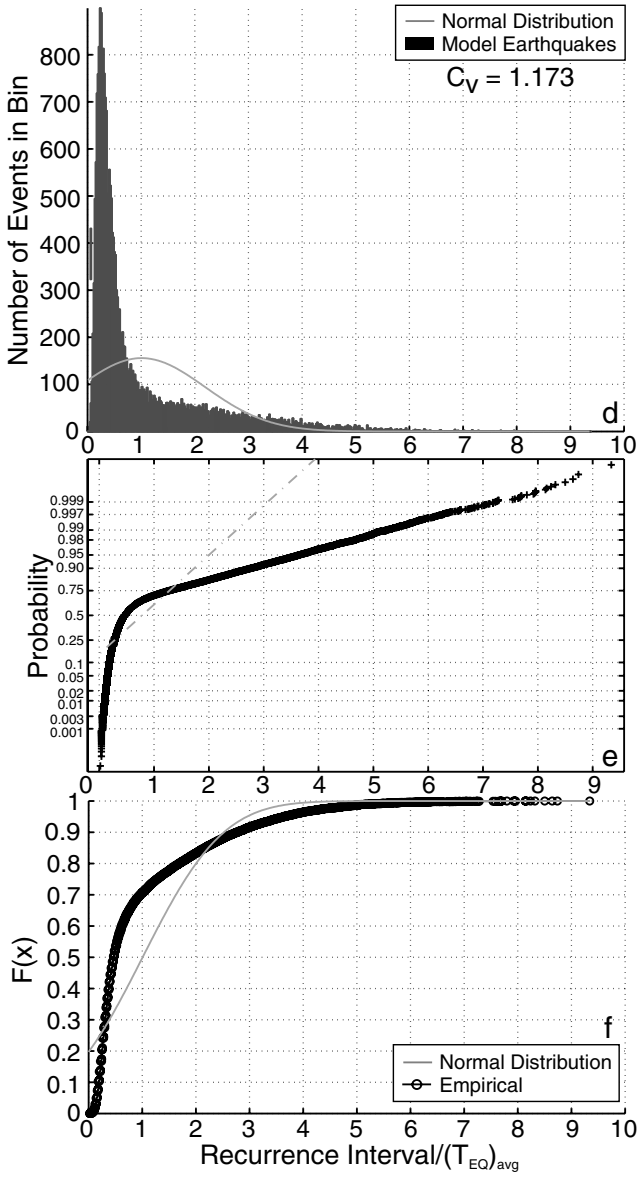

Figure 7. (*Continued.*)

and/or slip rates), a smaller proportion of the available post-seismic stress is returned to the crust during a typical inter-event interval. As a result, the strain-rate boundary condition represents the primary mechanism for increasing stress on the fault. Because the applied long-term rate is constant in time, the coefficient of variation calculated from output earthquake repeat times approaches that of the noise originally input into the system via the boundary condition. If the time required for complete relaxation of layers 2–3 is short relative to the average recurrence interval (i.e. high-W environments with low viscosity and/or slip rates), then a greater proportion of the available post-seismic stress is transferred to the crust before the next earthquake. Transferred post-seismic stresses therefore account for a larger percentage of the stress necessary to produce failure in the next event. If the amount of available post-seismic stress is variable, the result is a second time-dependent influence on earthquake recurrence intervals. Earthquake clustering may result.

In this model, temporal clustering of major earthquakes is induced by a combination of system noise (applied though τ^*_{max}) and post-seismic stress transfer. Any earthquake will transfer stress to the non-seismogenic layers of the model (e.g. Figs 8b–c and h–i). Relaxation of this post-seismic stress concentration generates crustal deformation rates that are initially higher than the geological average (e.g. Figs 8e and k). The larger the stress concentration is, the higher the resultant post-seismic deformation rate. For example, if an unusually large event, EQ_i, is randomly followed by an event, EQ_{i+1}, that is smaller than average, the second event, EQ_{i+1}, will occur sooner than expected because (i) the stress gain required to trigger EQ_{i+1} is smaller than average and (ii) as EQ_i was large, post-seismic fault loading rates will be higher than usual. Further, because EQ_{i+1} occurred sooner than expected, a smaller portion of the post-seismic stress concentration will have been recycled to the crust at the time of EQ_{i+1}. Therefore, as a result of random fluctuations in failure

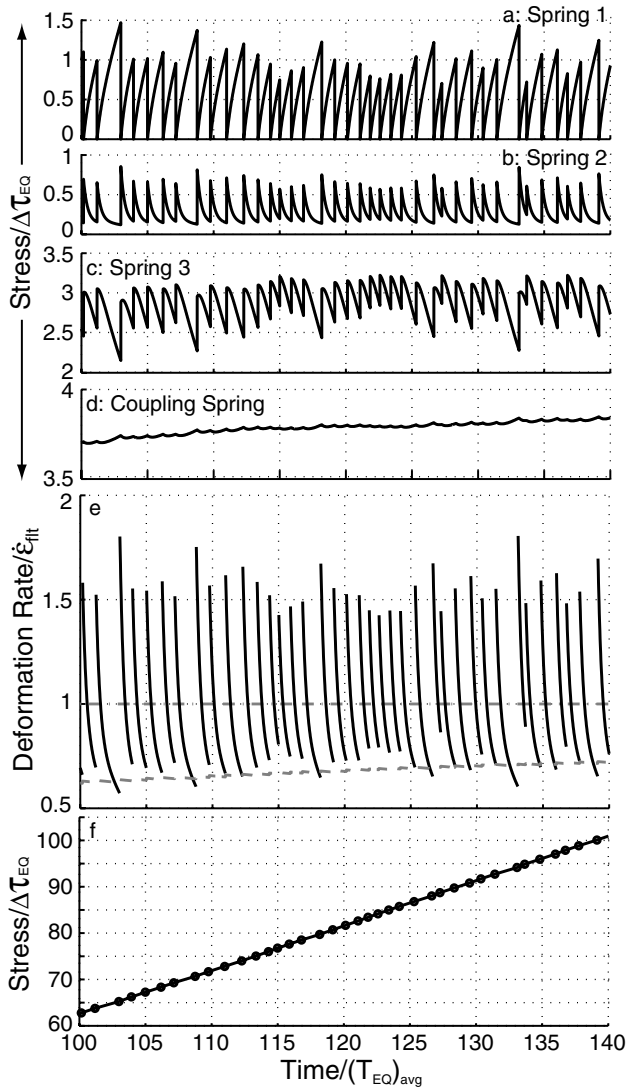

Figure 8. Temporal evolution of (a–d, g–j) stress in each element group, (e, k) crustal strain rate (solid) and (f, l) cumulative stress release for models with (a–f) $W = 2$ and (g–l) $W = 40$. In both cases, $\eta_3/\eta_2 = 10$, $G_c/G_1 = 0.1$, $G_2/G_1 = 1$, $G_3/G_1 = 1$ and $N_{max} = 20$ per cent. Note that in (a–d and g–j) the stress scales are not uniform. In (e, k), as time increases, the average fault strain release rate (dashed) asymptotically approaches the long-term geological rate (dash-dotted).

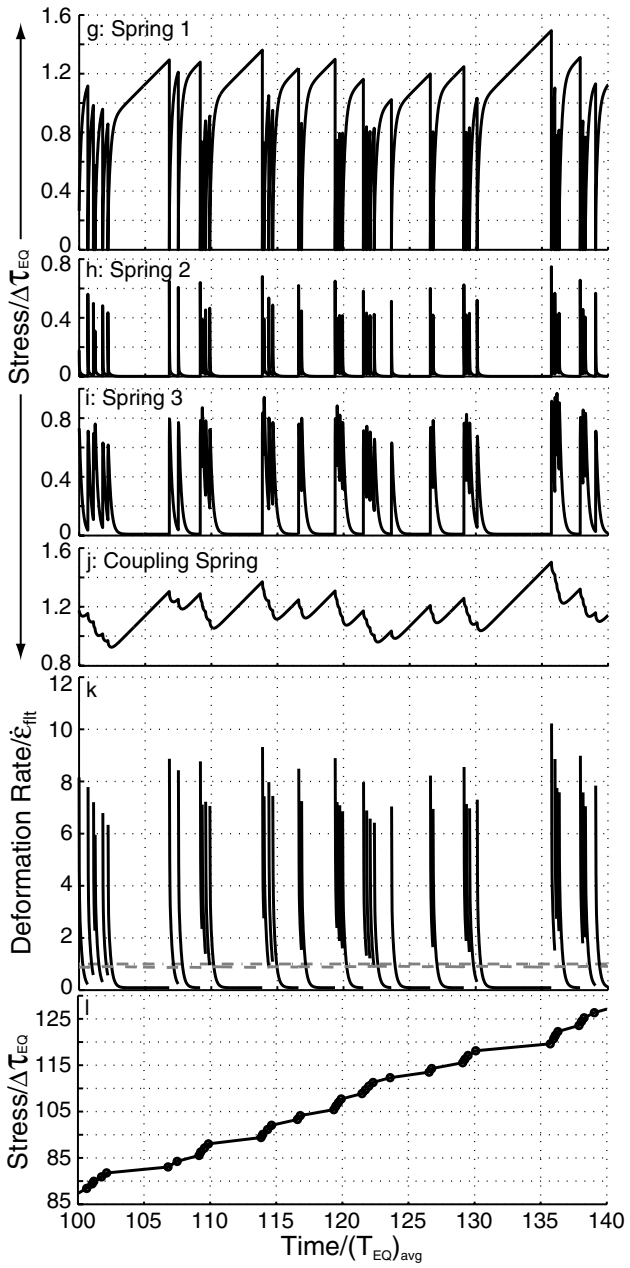

Figure 8. (*Continued.*)

stress/recurrence interval, the pool of available post-seismic stress in the non-seismogenic layers may become large compared to the available post-seismic stress following an isolated earthquake. Until this excess is exhausted, one would expect earthquakes to occur at a higher than average rate. It is likely, for example, that EQ_{i+2} will also occur earlier than expected.

Conversely, if an unusually small event is followed by an unusually large earthquake, one would expect a longer than average recurrence interval between the two events because the amount of post-seismic stress available for recycling to the seismogenic crust is smaller than average. More slowly accumulating tectonic loads must account for this deficit. This effect is especially pronounced if the pool of available post-seismic stress in the non-seismogenic layers is already diminished as a consequence of preceding earthquake sequences. One may therefore conclude that, if (i) the pool of available post-seismic stress in the non-seismogenic layers is variable in magnitude and (ii) post-seismic stresses comprise a significant percentage of the failure stress budget for a particular fault, random stress interactions between succeeding earthquakes on the same fault may lead to distinctly clustered behaviour.

Our findings regarding the sources of stress that drive faulting also have potential implications for the analysis of geodetic data. We note that in low-W environments where clustering does not play a significant role in the system behaviour (similar to the hypothetical plate boundary example, Figs 7a–c and 8a–f), peak post-seismic crustal strain rates are only ∼1.5 times greater than the applied geological rate (Fig. 8e). Reasonable choices of G_c/G_1 do not affect this result. Because high far-field plate velocities contribute the majority of the stress needed for fault failure, post-seismic transients are present but are of the same order of magnitude as the average geological rate. As stated previously, similar conditions are observed along the San Andreas fault (Thatcher 1983; Kenner & Segall 2000). Secondly, because earthquakes recur frequently, crustal deformation rates do

not return to their steady-state values before the next earthquake (Figs 8a–e). In other words, the system never reaches a true steady state and transients are always present. Consequently, a background interseismic strain rate is never attained.

Conversely, in clustered, high-W, low-strain-rate, less viscous environments (similar to the hypothetical Basin and Range example, Figs 7d–f and 8g–l), post-seismic transients (require $G_c/G_1 < 1$) contribute a significant portion of the stress required for failure in succeeding earthquakes. As a result, peak post-seismic strain rates immediately following an earthquake can be more than an order of magnitude greater than the geological average (Fig. 8k). These strain-rate transients can be greater than the geologic average for $\sim$15 per cent of the average time between earthquakes, $T_{\rm avg}^{\rm eq^*}$, such that, if $T_{\rm avg}^{\rm eq^*} \geq$ thousands of years, prolonged periods of anomalously high strain rate can result. At long times since the last earthquake, stresses and strain rates in every layer attain their steady-state values (Figs 8g–k) and all transients are absent. Thus, over the remainder of the earthquake cycle, initially high strain rates are counterbalanced by even longer intervals with strain rates well below the geologic average calculated using characteristic earthquake slip and recurrence interval data from historic and palaeoseismic records. This steady-state strain rate represents the background rate resulting from tectonic loading in the absence of all post-seismic transients. Removal of this background interseismic' rate to isolate post-seismic transients is reasonable assuming it can be accurately estimated. Average long-term geological rates, which may be much higher than the background steady-state rate, can not be used to isolate the post-seismic signal. Finally, the significant variations in deformation rate with time observed in high-W environments may help to explain anomalously high or low geodetic strain rates at various locations in the Basin and Range (e.g. Thatcher *et al.* 1999; Wernicke *et al.* 2000; Niemi *et al.* 2004).

In both high- and low-W environments, therefore, transient deformation complicates the comparison of instantaneous geodetic rates with geological estimates averaged over much longer time periods. This is especially true in low-strain-rate tectonic environments. For example, geodetic to geologic slip-rate comparisons have been found to be a problem in areas like the Basin and Range (Niemi *et al.* 2004; Friedrich *et al.* 2003). Taken in combination with major earthquake clustering, these temporal variations in deformation rate make the estimation of average repeat times for use in hazard estimation extremely difficult.

9 CONCLUSIONS

The spring-dashpot-slider model demonstrates that, over multiple event timescales, post-seismic processes can be an important factor in modulating recurrence intervals between major earthquakes on a single fault. Natural variations in earthquake repeat times resulting from environmental noise may create unusually large (small) concentrations of post-seismic stress in the non-seismogenic lithosphere. These variations in the magnitude of the post-seismic stress concentration at depth mean that post-seismic stressing rates along the coseismic fault are also variable. In tectonic environments where recycled post-seismic stresses represent a significant portion of the total stress required for fault failure and post-seismic stressing rates are unusually high (low), the next earthquake will occur sooner (later) than expected. Over multiple event sequences, temporal clustering of large earthquakes may result.

Earthquake clustering behaviour in different tectonic environments is a function of the non-dimensional Wallace number, W. As W increases, two distinct behavioural regimes are observed. They are separated by a clear transition zone (Figs 4–6). This clustered versus non-clustered behaviour is similar to the bimodal behaviour observed by Ben-Zion *et al.* (1999) and Lyakhovsky *et al.* (2001) using continuum damage model rheologies. Such a damage rheology may, therefore, contribute to the environmental noise employed in this study. We find that faults in high-W tectonic regimes (low long-term slip-rates in regions with low viscosities in the non-seismogenic lithosphere and high stress drop earthquakes) are more susceptible to earthquake clustering resulting from post-seismic stress transfer. The first two characteristics are synonymous with regions like the Basin and Range. Individual faults within the Basin and Range experience very low long-term slip-rates (e.g. Niemi *et al.* 2004), while the region as a whole is known to have relatively thin effective elastic thicknesses and high heat flow (Lowry & Smith 1995). In particular, the Wasatch fault slips at $\sim$ 2–3 mm yr^{-1}. Lesser faults scattered throughout the interior of the Basin and Range slip at rates much less than 1 mm yr^{-1}. In these hot, weak, low-strain-rate environments, the non-seismogenic lithosphere effectively recycles post-seismic stress to the seismogenic crust and far-field tectonic strain rates are low enough that, over periods of a few earthquake cycles, unusually large post-seismic transients can have a significant affect on earthquake recurrence intervals. As a result, in regions like the Basin and Range, temporal clustering or swarms of major earthquakes on a single fault are an expected behaviour.

Secondly, we show that, in low-W tectonic environments, possibly similar to those along the major plate boundaries such as the San Andreas fault, crustal deformation rates may never attain a steady-state background value. Rates, may, in fact, vary continuously. In high-W environments like the Basin and Range with average earthquake recurrence intervals that are long compared with the effective lithospheric relaxation time, steady-state rates well below the long-term geologic average may be attained. These low rates are coupled with extremely high peak strain rates that remain well above the geological average for $\sim$ 15 per cent of the earthquake cycle. These temporal variations mean that the interpretation of geodetic data for hazard estimation is potentially dependent on the recent seismic history of the fault. Discrepancies in fault slip-rates inferred from geodetic and geologic measurements made over different time intervals are a direct consequence.

Finally, the presented spring-dashpot-slider model demonstrates that modelled crustal deformation rates and lithospheric stresses evolve throughout an initial model cycle-up process, which includes two stages. These stages may be superposed during the actual cycle-up process. The first stage represents steady-state conditions in the absence of earthquakes. The second steady state is attained following the initiation of repeating earthquakes. This initial temporal evolution in stress and strain rate highlights the potential importance of preconditioning stresses in more complex multidimensional numerical models of geodetic data.

ACKNOWLEDGMENTS

We would like to thank Luc Lavier for the many helpful discussions he participated in while this work was in progress. This work was partially funded by NSF grant EAR-0229793.

REFERENCES

Al-Zoubi, A. & ten Brink, U., 2002. Lower crustal flow and the role of shear in basin subsidence: An example from the Dead Sea basin, *Earth planet. Sci. Lett.,* **199,** 67–79.

Anderson, J.G., 1986. Seismic strain rates in the central and eastern United States, *Bull. seism. Soc. Am.,* **76,** 273–290.

Ben-Zion, Y. & Rice, J.R., 1995. Slip patterns and earthquake populations along different classes of faults in elastic solids, *J. geophys. Res.,* **100,** 12 959–12 983.

Ben-Zion, Y., Dahmen, K., Lyakhovsky, V., Ertas, D. & Agnon, A., 1999. Self-driven mode switching of earthquake activity on a fault system, *Earth planet. Sci. Lett.,* **172,** 11–21.

Burridge, R. & Knopoff, L., 1967. Model and theoretical seismicity, *Bull. seism. Soc. Am.,* **57,** 341–371.

Carlson, J.M. & Langer, J.S., 1989. Mechanical model of a earthquake fault, *Phys. Rev. A,* **40,** 6470–6484.

Chéry, J., Merkel S. & Bouissou, S., 2001. A physical basis for time clustering of large earthquakes, *Bull. seism. Soc. Am.,* **91,** 1685–1693.

Chéry, J., Carretier, S. & Ritz, J-F, 2001. Postseismic stress transfer explains time clustering of large earthquakes in Mongolia, *Earth planet. Sci. Lett.,* **194,** 277–286.

Dieterich, J., 1994. A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. geophys. Res.,* **99,** 2601–2618.

Dorsey, R.J., Umhoefer, P.J. & Falk P.D., 1997. Earthquake clustering inferred from Pliocene Gilbert-type fan deltas in the Loreto basin, Baja California Sur, Mexico, *Geology*, **25,** 679–682.

Friedrich, A.M.Wernicke, B.P., Niemi, N.A., Bennet, R.A. & Davis, J.L., 2003. Comparison of geodetic and geologic data from the Wasatch region, Utah and implications for the spectral character of earth deformation at periods of 10 to 10 million years, *J. geophys. Res.,* **108,** 2199, doi:10.1029/2001JB000682.

Goes, S.D.B., 1996. Irregular recurrence of large earthquake: An analysis of historic and Paleoseismic catalogs, *J. geophys. Res.,* **101,** 5739–5749.

Grant, L.B. & Sieh, K., 1994. Paleoseismic evidence of clustered earthquakes on the San Andreas fault in the Carrizo Plain, California, *J. geophys. Res.,* **99,** 6819–6841.

Hager, B.H., Lyzenga, G.A., Donnellan, A. & Dong, D., 1999. Reconciling rapid strain accumulation with deep seismogenic fault planes in the Ventura basin, California, *J. geophys. Res.,* **104,** 25 207–25 219.

Hainzl, S., Zöller, G. & Kurths, J., 1999. Similar power laws for foreshock and aftershock sequences in a spring-block model for earthquakes, *J. geophys. Res.,* **104,** 7243–7253.

Hainzl, S., Zöller, G. & Kurths, J., 2000. Self-organization of spatio-temporal earthquake clusters, *Nonlinear Processes Geophys.*, **7,** 21–29.

Harris, R.A. & Simpson, R.W., 1998. Suppression of large earthquakes by stress shadows; a comparison of Coulomb and rate-and-state failure, *J. geophys. Res.,* **103,** 24 439–24 451.

Kagan, Y.Y. & Jackson, D.D., 1991. Long-term earthquake clustering, *Geophys. J. Int.,* **104,** 117–133.

Kagan, Y.Y. & Knopoff, L., 1976. Statistical search for non-random features of the seismicity of strong earthquakes, *Phys. Earth planet. Int.,* **12,** 291–318.

Kenner, S.J., 2004. Rheological controls on fault loading rates in northern California following the 1906 San Francisco earthquake, *Geophys. Res. Lett.,* **31,** L01606, doi:10.1029/2003GL018903.

Kenner, S.J. & Segall, P., 1999. Time-dependence of the Stress Shadowing Effect and Its Relation to the Structure of the Lower Crust, *Geology,* **27,** 119–122.

Kenner, S.J. & Segall, P., 2000. Postseismic Deformation Following the 1906 San Francisco Earthquake, *J. geophys. Res.,* **105,** 13 195–13 209.

King, G.C.P., Stein, R.S. & Lin, J., 1994. Static stress changes and the triggering of earthquakes, *Bull. seism. Soc. Am.,* **84,** 935–953.

Klinger, Y., Avouac, J.P., Abou Karaki, N., Dorbath, L., Bourles, D. & Reyss J.L., 2000. Slip rate on the Dead Sea transform fault in northern Araba valley (Jordan), *Geophys. J. Int.,* **142,** 755–768.

Lehner, F.K. & Li, V.C., 1982. Large-scale characteristics of plate boundary deformations related to the post-seismic readjustment of a thin lithosphere, *Geophys. J. R. astr. Soc.,* **71,** 775–792.

Li, V.C. & Kisslinger, C., 1985. Stress transfer and nonlinear stress accumulation at subduction-type plate boundaries: Application to the Aleutians, *Pure appl. Geophys.,* **122,** 812–830.

Li, V.C. & Rice, J.R., 1987. Crustal deformation in great California earthquake cycles, *J. geophys. Res.,* **92,** 11 533–11 551.

Lowry, A.R. & Smith, R.B., 1995. Strength and rheology of the western U.S. cordillera, *J. geophys. Res.,* **100,** 17 947–17 963.

Lyakhovsky, V., Ben-Zion, Y. & Agnon, A., 2001. Earthquake cycle, fault zones, and seismicity patterns in a rheologically layered lithosphere, *J. geophys. Res.,* **106,** 4103–4120.

Lynch, J.C. & Richards, M.A., 2001. Finite element models of stress orientations in well-developed strike-slip fault zones: Implications for the distribution of lower crustal strain, *J. geophys. Res.,* **106,** 26 707–26 729.

Lynch, J.C., Bürgmann, R., Richards, M.A. & Ferencz, R.M., 2003. When faults communicate: Viscoelastic coupling and earthquake clustering in a simple two-fault system, *Geophys. Res. Lett.,* **30,** 1270, doi:10.1029/2002GL016765.

Lyzenga, G.A., Raefsky, A. & Mulligan, S.G., 1991. Models of recurrent strike-slip earthquake cycles and the state of crustal stress, *J. geophys. Res.,* **96,** 21 623–21 640.

McCalpin, J.P. & Nishenko, S.P., 1996. Holocene paleoseismicity, temporal clustering, and probabilities of future large ($M > 7$) earthquakes on the Wasatch fault zone, Utah, *J. geophys. Res.,* **101,** 6233–6253.

Marco, S., Stein, M. & Agnon, A., 1996. Long-term earthquake clustering: A 50 000-year paleoseismic record in the Dead Sea Graben, *J. geophys. Res.,* **101,** 6179–6191.

Matthews, M.V., Ellsworth, W.L. & Reasonberg, P.A., 2002. A Brownian Model for Recurrent Earthquakes, *Bull. seism. Soc. Am.,* **92,** 2233–2250.

Niemi, N.A., Wernicke, B.P., Friedrich, A.M., Bennett, R.A. & Davis, J.L., 2004. BARGEN continuous GPS data across the eastern Basin and Range province and implications for fault system dynamics, *Geophys. J. Int.,* **159** in press.

Nomanbhoy, N. & Ruff, L.J., 1996. A simple discrete model for large multiplet earthquakes, *J. geophys. Res.,* **101,** 5707–5723.

Pelletier, J.D., 2000. Spring-block models of seismicity: Review and analysis of a structurally heterogeneous model coupled to a viscous asthenosphere, in *Geocomplexity and the Physics of Earthquakes,* Geophysical Monograph 120, pp. 27–42, eds Rundle, J.B. Turcotte, D.L. & Klein, W., American Geophysical Union, Washington, DC.

Pratt, T.L., 1994. How old is the New Madrid seismic zone?, *Seismol. Res. Lett.,* **65,** 172–179.

Reches, Z., Schubert, G. & Anderson, C., 1994. Modeling of periodic great earthquakes on the San Andreas fault: Effects of nonlinear crustal rheology, *J. geophys. Res.,* **99,** 21 983–22 000.

Ritz, J.F., Brown, E.T., Bourlès, D.L., Philip, H., Schlupp, A., Raisbeck, G.M., Yiou, F. & Enkhtuvshin, B., 1995. Slip rates along active faults estimated with cosmic-ray-exposure dates: Application to the Bogd fault, Gobi-Altï, Mongolia, *Geology,* **23,** 1019–1022.

Rockwell, T.K., Lindvall, S., Herzberg, M., Murbach, D., Dawson, T. & Berger, G., 2000. Paleoseismology of the Johnson Valley, Kickapoo, and Homestead Valley faults: Clustering of earthquakes in the Eastern California Shear Zone, *Bull. seism. Soc. Am.,* **90,** 1200–1236.

Rundle, J.B., 1986. An approach to modeling present-day deformation in southern California, *J. geophys. Res.,* **91,** 1947–1959.

Rundle, J.B., 1988. A physical model for earthquakes; 2, Application to Southern California, *J. geophys. Res.,* **93,** 6255–6274.

Savage, J.C. & Plafker, G., 1991. Tide gage measurements of uplift along the south coast of Alaska, *J. geophys. Res.,* **96,** 4325–4335.

Schweig, E.S. & Ellis, M.A., 1994. Reconciling short recurrence intervals with minor deformation in the New Madrid seismic zone, *Science,* **264,** 1308–1311.

Stein, R.S., Barka, A.A. & Dieterich, J.H., 1997. Progressive failure of the North Anatolian fault since 1939 by earthquake stress triggering, *Geophys. J. Int.,* **128,** 594–604.

Swan, F.H., 1988. Temporal clustering of Paleoseismic events on the Oued Fodda fault, Algeria, *Geology,* **16,** 1092–1095.

Thatcher, W., 1983. Nonlinear strain buildup and the earthquake cycle on the San Andreas fault, *J. geophys. Res.,* **88,** 5893–5902.

Thatcher, W., 1984. The earthquake deformation cycle at the Nankai Trough, southwest Japan, *J. geophys. Res.,* **87,** 3087–3101.

Thatcher, W., 1990. Present-day crustal movements and the mechanics of the deformation cycle, in, *The San Andreas fault system, California,* Professional Paper 1515, pp. 189–205, ed. Wallace, R.E., US Geological Survey, Denver, CO.

Thatcher, W., Foulger, G.R., Julian, B.R., Svarc, J., Quilty, E. & Bawden, G.W., 1999. Present-day deformation across the Basin and Range Province, Western United States, *Science,* **283,** 1714–1718.

Turcotte, D.L., Newman, W.I. & Gabrielov, A., 2000. A statistical approach to earthquakes, in, *Geocomplexity and the Physics of Earthquakes,*Geophysical Monograph 120, pp. 83–96, eds Rundle, J.B., Turcotte, D.L. & Klein, W., American Geophysical Union, Washington, DC.

Tuttle, M.P., Schweig, E.S., Sims, J.D., Lafferty, R.H., Wolf, L.W. & Haynes, M.L., 2002. The earthquake potential of the New Madrid Seismic Zone, *Bull. seism. Soc. Am.,* **92,** 2080–2089.

Van Arsdale, R., 2000. Displacement history and slip rate on the Reelfoot fault of the New Madrid seismic zone, *Engineering Geology,* **55,** 219–226.

Wallace, R.E., 1987. Grouping and migration of surface faulting and variations in slip rates on faults in the Great Basin province, *Bull. seism. Soc. Am.,* **77,** 868–876.

Wernicke, B.P., Friedrich, A.M., Niemi, N.A., Bennett, R.A. & Davis, J.L., 2000. Dynamics of plate boundary fault systems from Basin and Range Geodetic Network (BARGEN) and geologic data, *GSA Today,* **10,** 1–7.

Xu, X. & Deng, Q., 1996. Nonlinear characteristics of paleoseismicity in China, *J. geophys. Res.,* **101,** 6209–6231.

APPENDIX A: MATERIAL PARAMETERS AND NON-DIMENSIONALIZATION

The constants in eq. (1), the dimensional interseismic governing equations, are defined as follows:

$$\begin{aligned}
\delta_1 &= G_c\eta_2,\\
\delta_0 &= G_2G_c,\\
\gamma_1 &= \eta_2(G_1+G_2+G_c),\\
\gamma_0 &= G_2(G_1+G_c),\\
\alpha_2 &= G_c\eta_2\eta_3,\\
\alpha_1 &= G_c(\eta_2G_3+\eta_3G_2),\\
\alpha_0 &= G_2G_3G_c,\\
\beta_2 &= \eta_2\eta_3(G_1+G_2+G_3+G_c),\\
\beta_1 &= (\eta_2G_3+\eta_3G_2)(G_1+G_c)+G_2G_3(\eta_2+\eta_3),\\
\beta_0 &= G_2G_3(G_1+G_c).
\end{aligned} \tag{A1}$$

Employing the non-dimensionalization described in eq. (3), we find that the material constants G_i and η_i in the simplified governing equations always appear as the ratios G_i/G_1 and η_i/η_{eff}. Given this approach, the governing equations for a given spring and dashpot j become, $\tau^*_{Sj}=(G_j/G_1)\varepsilon^*_{Sj}$ and $\tau^*_{Dj}=2(\eta_j/\eta_{\text{eff}})(1/W)\dot{\varepsilon}^*_{Dj}$, respectively.

Using eq. (3), we find that the governing equations are also automatically normalized. After non-dimensionalizing $T^{\text{eq}}_{\text{avg}}=\Delta\varepsilon_{\text{eq}}/\dot{\varepsilon}_{\text{flt}}$, we find that $T^{\text{eq}^*}_{\text{avg}}=\Delta\varepsilon^*_{\text{eq}}$. Thus, when eq. (4a) is used, all non-dimensional times and strains in the solutions are found to be divided by $T^{\text{eq}^*}_{\text{avg}}$ and $\Delta\varepsilon^*_{\text{eq}}$, respectively. In consequence, the non-dimensional, normalized time and strain are defined as $t^{*n}=t^*/T^{\text{eq}*}_{\text{avg}}=(W\dot{\varepsilon}_{\text{flt}}t)/(T^{\text{eq}*}_{\text{avg}})=(W\dot{\varepsilon}_{\text{flt}}t)/(\Delta\varepsilon^*_{\text{eq}})$ and $\varepsilon^{*n}=\varepsilon^*/\Delta\varepsilon^*_{\text{eq}}=W\varepsilon/\Delta\varepsilon^*_{\text{eq}}$, where $\Delta\varepsilon^*_{\text{eq}}=W\Delta\varepsilon_{\text{eq}}$. Based on these definitions, it can also be shown that $\dot{\varepsilon}^{*n}=\dot{\varepsilon}^*$.

APPENDIX B: TWO-LAYER SOLUTION

B1 Governing equations

After normalization and non-dimensionalization, the two-layer governing eq. (2) has the form

$$P_1\dot{\varepsilon}^*_{\text{lyr}}+WP_0\varepsilon^*_{\text{lyr}}=W\left[1+Q_0\left(t^*+\varepsilon^*_{\text{tot}}\big|_{t=0}\right)\right], \tag{B1}$$

where P_i and Q_i are non-dimensional functions of the material parameters obtained by dividing through by δ_1 and non-dimensionalizing the result using eq. (4a). For example, $Q_0=(\delta_0/\delta_1)/(W\dot{\varepsilon}_{\text{flt}})$ because δ_0/δ_1 has units of 1/time. Also note that the non-dimensionalized boundary condition becomes $\dot{\varepsilon}^*_{\text{flt}}=1$.

B2 Solution

The non-dimensional solution to the two-layer problem is given by eq. (5). Before specifying Γ and Π_i, define

$$\begin{aligned}
\delta^n_1 &= G^n_c\eta^n_2,\\
\delta^n_0 &= G^n_2G^n_c,\\
\gamma^n_1 &= \eta^n_2\left(1+G^n_2+G^n_c\right),\\
\gamma^n_0 &= G^n_2\left(1+G^n_c\right),
\end{aligned} \tag{B2}$$

where $G^n_i=G_i/\,G_1$ and $\eta^n_i=\eta_i/\,\eta_{\text{eff}}$. Using $G_1=\Delta\tau_{\text{eq}}/\,2\Delta\varepsilon_{\text{eq}}$ to normalize terms containing t in eq. (5), Γ and Π_i can then be written as

$$\begin{aligned}
\Gamma &= \frac{2}{W}\frac{\gamma_1^n}{\gamma_0^n},\\
\Pi_1 &= \frac{2}{W}\left(\frac{\delta_1^n}{\gamma_0^n} - \frac{\delta_0^n\gamma_1^n}{\left(\gamma_0^n\right)^2}\right) + \frac{\delta_0^n}{\gamma_0^n}\varepsilon^*_{\rm tot}\big|_{t^*=0},\\
\Pi_2 &= \varepsilon^*_{\rm lyr}\big|_{t^*=0},\\
\Pi_3 &= \frac{\delta_0^n}{\gamma_0^n}
\end{aligned} \tag{B3}$$

Note that all strains and times in eq. (B3) have been normalized (Appendix A).

B3 Determining the stress/strain in each element

The strain in each element can be found using the following equations:

$$\begin{aligned}
\varepsilon^*_{S1}(t^*) &= \varepsilon^*_{\rm lyr}(t^*),\\
\varepsilon^*_{S2}(t^*) &= -\Pi_1\frac{R}{D_2+R}\left[\exp\left(\frac{RWt^*}{2}\right) - \exp\left(\frac{-D_2Wt^*}{2}\right)\right]\\
&\quad + \frac{\Pi_2}{D_2+R}\left[R\exp\left(\frac{RWt^*}{2}\right) + D_2\exp\left(\frac{-D_2Wt^*}{2}\right)\right]\\
&\quad + \frac{2\Pi_3}{D_2W}\left[1-\exp\left(\frac{-D_2Wt^*}{2}\right)\right] - \varepsilon_{D2}|_{t^*=0}\exp\left(\frac{-D_2Wt^*}{2}\right),\\
\varepsilon^*_{D2}(t^*) &= \Pi_1\left[1-\exp\left(\frac{-D_2Wt^*}{2}\right)\right] + \frac{D_2}{D_2+R}\left[\Pi_2-\Pi_1\right]\left[\exp\left(\frac{RWt^*}{2}\right) - \exp\left(\frac{-D_2Wt^*}{2}\right)\right]\\
&\quad - \frac{2\Pi_3}{D_2W}\left[1-\exp\left(\frac{-D_2Wt^*}{2}\right)\right] + \Pi_3t^* + \varepsilon_{D2}|_{t^*=0}\exp\left(\frac{-D_2Wt^*}{2}\right),\\
\varepsilon^*_{Sc}(t^*) &= \left(\varepsilon^*_{\rm tot}\big|_{t^*=0} + \dot\varepsilon^*_{\rm flt}t^*\right) - \varepsilon^*_{\rm lyr}(t^*),
\end{aligned} \tag{B4}$$

where $D_2 = G_2^n/\,\eta_2^n$, $R = -\,\gamma_0^n/\,\gamma_1^n = -\,2/\,\Gamma W$ and all times have been normalized. The stress in any element group $j = 1, 2$, or c can be derived using $\tau_j^* = (G_j/G_1)\varepsilon^*_{Sj}$. Note that initial dashpot displacement must also be specified. $\varepsilon^*_{\rm lyr}(0)$, $\varepsilon^*_{\rm tot}(0)$ and the initial dashpot strain are defined below.

B4 Resetting the system following a slip event

At the time of an earthquake, $t^*_{\rm eq}$, stresses in the two-layer model evolve as follows:

$$\begin{aligned}
\tau_1^*\left(t_{\rm eq}^{*+}\right) &= \tau^*_{\rm residual},\\
\tau_2^*\left(t_{\rm eq}^{*+}\right) &= \tau_2^*\left(t_{\rm eq}^{*-}\right) + \left(\tau^*_{\rm fail} - \tau^*_{\rm residual}\right),\\
\tau_c^*\left(t_{\rm eq}^{*+}\right) &= \tau_c^*\left(t_{\rm eq}^{*-}\right).
\end{aligned} \tag{B5}$$

Following an earthquake, the time since the last event, t^*, is reset to zero. As shown below, other initial conditions are reset such that all dimensions in the system, neglecting cumulative slider displacements, are internally consistent:

$$\begin{aligned}
\varepsilon^*_{\rm lyr}|_{t^*=0} &= \tau^*_{\rm residual},\\
\varepsilon^*_{\rm tot}|_{t^*=0} &= \varepsilon^*_{\rm lyr}|_{t^*=0} + \frac{\tau_c^*\left(t_{\rm eq}^{*+}\right)}{G_c^n},\\
\varepsilon^*_{Dj}|_{t^*=0} &= \tau^*_{\rm residual} - \frac{\tau_j^*\left(t_{\rm eq}^{*+}\right)}{G_j^n}.
\end{aligned} \tag{B6}$$

APPENDIX C: THREE-LAYER SOLUTION

C1 Governing equations

For a constant strain-rate boundary condition, $\dot\varepsilon_{\rm tot} = \dot\varepsilon_{\rm flt}$, eq. (1b) reduces to

$$\beta_2\ddot\varepsilon_{\rm lyr} + \beta_1\dot\varepsilon_{\rm lyr} + \beta_0\varepsilon_{\rm lyr} = (\alpha_0t + \alpha_1)\dot\varepsilon_{\rm flt} + \alpha_0\ \varepsilon_{\rm tot}|_{t=0}\,, \tag{C1}$$

where the constant velocity boundary condition, $V_{\rm flt}$, has been reformulated in terms of strain rate to maintain dimensional consistency. After normalization and non-dimensionalization, this becomes

$$S_2\ddot{\varepsilon}^*_{\rm lyr} + W\left(S_1\dot{\varepsilon}^*_{\rm lyr} + S_0\varepsilon^*_{\rm lyr}\right) = W\left[M_1 + M_0\left(t^* + \varepsilon^*_{\rm tot}\big|_{t=0}\right)\right], \tag{C2}$$

where S_i and M_i are non-dimensional functions of the material parameters obtained by dividing through by α_1 and non-dimensionalizing the result using eq. (4a) (see Appendix B).

C2 Solution

The three-layer problem (C2) is a second order differential equation. The non-dimensional solution to eq. (C2) has the form

$$\varepsilon^*_{\rm lyr} = \Omega_1 \exp[-\Phi_3(\Phi_1 - \Phi_2)t^*] + \Omega_2 \exp[-\Phi_3(\Phi_1 + \Phi_2)t^*] + \Omega_3(\varepsilon^*_{\rm tot}|_{t=0} + t^*) + \Omega_4, \tag{C3}$$

where Ω_i and Φ_i are non-dimensional functions of the material parameters, W, $\varepsilon^*_{\rm tot}$ (0), $\varepsilon^*_{\rm lyr}$ (0) and $\dot{\varepsilon}^*_{\rm lyr}(0)$ as shown below. Before specifying Ω_i and Φ_i, define

$$\begin{aligned}
\alpha_2^n &= G_c^n\eta_2^n\eta_3^n,\\
\alpha_1^n &= G_c^n\left(\eta_2^n G_3^n + \eta_3^n G_2^n\right),\\
\alpha_0^n &= G_2^n G_3^n G_c^n,\\
\beta_2^n &= \eta_2^n\eta_3^n\left(1 + G_2^n + G_3^n + G_c^n\right),\\
\beta_1^n &= \left(\eta_2^n G_3^n + \eta_3^n G_2^n\right)\left(1 + G_c^n\right) + G_2^n G_3^n\left(\eta_2^n + \eta_3^n\right),\\
\beta_0^n &= G_2^n G_3^n\left(1 + G_c^n\right),
\end{aligned} \tag{C4}$$

where $G_i^n = G_i/\,G_1$ and $\eta_i^n = \eta_i/\,\eta_{\rm eff}$. Using $G_1 = \Delta\tau_{\rm eq}/2\Delta\varepsilon_{\rm eq}$ to normalize terms containing t, Ω_i and Φ_i can then be written as

$$\begin{aligned}
\Omega_1 &= \Psi_2 + \Psi_1\left[\Psi_2 + \frac{2}{W}\left(-\frac{\alpha_0^n\beta_2^n}{\beta_0^n\beta_1^n} + \frac{\beta_2^n}{\beta_1^n}\,\dot{\varepsilon}^*_{\rm lyr}\big|_{t^*=0}\right)\right],\\
\Omega_2 &= \Psi_2 - \Psi_1\left[\Psi_2 + \frac{2}{W}\left(-\frac{\alpha_0^n\beta_2^n}{\beta_0^n\beta_1^n} + \frac{\beta_2^n}{\beta_1^n}\,\dot{\varepsilon}^*_{\rm lyr}\big|_{t^*=0}\right)\right],\\
\Omega_3 &= \frac{\alpha_0^n}{\beta_0^n},\\
\Omega_4 &= \frac{2}{W}\left(\frac{\alpha_1^n}{\beta_0^n} - \frac{\alpha_0^n\beta_1^n}{(\beta_0^n)^2}\right),\\
\Phi_1 &= \frac{\beta_1^n}{2\beta_2^n},\\
\Phi_2 &= \frac{\sqrt{\left(\beta_1^n\right)^2 - 4\beta_2^n\beta_0^n}}{2\beta_2^n},\\
\Phi_3 &= W/2
\end{aligned} \tag{C5}$$

where

$$\begin{aligned}
\Psi_1 &= \frac{\beta_1^n}{\sqrt{\left(\beta_1^n\right)^2 - 4\beta_2^n\beta_0^n}},\\
\Psi_2 &= \frac{2}{W}\left(\frac{-\alpha_1^n}{2\beta_0^n} + \frac{\alpha_0^n\beta_1^n}{2\left(\beta_0^n\right)^2}\right) - \frac{\alpha_0^n}{2\beta_0^n}\,\varepsilon^*_{\rm tot}\big|_{t^*=0} + \frac{1}{2}\,\varepsilon^*_{\rm lyr}\big|_{t^*=0}.
\end{aligned} \tag{C6}$$

Note that all strains and times in eqs (C5) and (C6) have been normalized (Appendix A). $R_1 = -\left(\Phi_1 - \Phi_2\right)$ and $R_2 = -\left(\Phi_1 + \Phi_2\right)$ are derived from the roots of the characteristic equation required to solve the differential eq. (C2). We presume that both R_1 and R_2 are real and non-zero, as is generally the case for reasonable definitions of lithospheric material properties.

C3 Determining the stress/strain in each element

Between earthquakes, the strain in each element can be found using the following equations:

$$
\begin{aligned}
\varepsilon^*_{S1}(t^*) &= \varepsilon^*_{\rm lyr}(t^*),\\
\varepsilon^*_{Sj}(t^*) &= \frac{\Omega_1}{D_j+R_1}\left[R_1\exp\left(\frac{R_1Wt^*}{2}\right)+D_j\exp\left(\frac{-D_jWt^*}{2}\right)\right]\\
&\quad+\frac{\Omega_2}{D_j+R_2}\left[R_2\exp\left(\frac{R_2Wt^*}{2}\right)+D_j\exp\left(\frac{-D_jWt^*}{2}\right)\right]\\
&\quad+\left(\Omega_3\ \varepsilon^*_{\rm tot}\big|_{t^*=0}+\Omega_4\right)\exp\left(\frac{-D_jWt^*}{2}\right)+\frac{2\Omega_3}{D_jW}\left[1-\exp\left(\frac{-D_jWt^*}{2}\right)\right]\\
&\quad-\varepsilon_{Dj}\big|_{t^*=0}\exp\left(\frac{-D_jWt^*}{2}\right),\\
\varepsilon^*_{Dj}(t^*) &= \frac{\Omega_1D_j}{D_j+R_1}\left[\exp\left(\frac{R_1Wt^*}{2}\right)-\exp\left(\frac{-D_jWt^*}{2}\right)\right]+\frac{\Omega_2D_j}{D_j+R_2}\left[\exp\left(\frac{R_2Wt^*}{2}\right)-\exp\left(\frac{-D_jWt^*}{2}\right)\right]\\
&\quad+\left(\Omega_3\ \varepsilon^*_{\rm tot}\big|_{t^*=0}+\Omega_4\right)\left[1-\exp\left(\frac{-D_jWt^*}{2}\right)\right]+\frac{2\Omega_3}{D_jW}\left[-1+\left(\frac{D_jW}{2}\right)t^*+\exp\left(\frac{-D_jWt^*}{2}\right)\right]\\
&\quad+\ \varepsilon_{Dj}\big|_{t^*=0}\exp\left(\frac{-D_jWt^*}{2}\right),\\
\varepsilon^*_{Sc}(t^*) &= \left(\varepsilon^*_{\rm tot}\big|_{t^*=0}+\dot\varepsilon^*_{\rm flt}t^*\right)-\varepsilon^*_{\rm lyr}(t^*),
\end{aligned}
\tag{C7}
$$

where $D_j = G^n_j/\eta^n_j$ for $j = 2, 3$. R_j is defined above and all times have been normalized. The stress in any element group $j = 1, 2, 3$, or c can be derived using $\tau^*_j = (G_j/G_1)\varepsilon^*_{Sj}$. Note that initial dashpot displacements must also be specified.

C4 Resetting the system following a slip event

During an earthquake, stress is distributed between the two viscoelastic layers in proportion to their elastic stiffness. Thus, immediately following an earthquake, the stresses in each element are defined as

$$
\begin{aligned}
\tau^*_1\left(t^{*+}_{\rm eq}\right) &= \tau^*_{\rm residual},\\
\tau^*_2\left(t^{*+}_{\rm eq}\right) &= \tau^*_2\left(t^{*-}_{\rm eq}\right)+\frac{G^n_2}{G^n_2+G^n_3}\left(\tau^*_{\rm fail}-\tau^*_{\rm residual}\right),\\
\tau^*_3\left(t^{*+}_{\rm eq}\right) &= \tau^*_3\left(t^{*-}_{\rm eq}\right)+\frac{G^n_3}{G^n_2+G^n_3}\left(\tau^*_{\rm fail}-\tau^*_{\rm residual}\right),\\
\tau^*_c\left(t^{*+}_{\rm eq}\right) &= \tau^*_c\left(t^{*-}_{\rm eq}\right)
\end{aligned}
\tag{C8}
$$

Required initial strains are reset according using the definitions given in eq. (B6) for the two-layer model. Conservation of stress within the system requires that, when summed, the rates of change of stress in elements 1–3 must equal the rate of change of stress in the coupling spring. Using this constraint, we can determine the remaining initial strain rate:

$$
\dot\varepsilon^*_{\rm lyr}\big|_{t^*=0}=\frac{G^n_c+0.5W\left(G^n_2D_2+G^n_3D_3\right)\varepsilon^*_{\rm lyr}\big|_{t^*=0}-0.5WG^n_2D_2\ \varepsilon^*_{D2}\big|_{t^*=0}-0.5WG^n_3D_3\ \varepsilon^*_{D3}\big|_{t^*=0}}{G^n_1+G^n_2+G^n_3+G^n_c}.
\tag{C9}
$$

APPENDIX D: NORMALLY DISTRIBUTED NOISE DISTRIBUTION

As stated in the text, the noise on the fault failure criterion is normally distributed and $N_{\rm residual}$ is fixed with zero magnitude. Because a normal distribution allows negative values for large $N_{\rm max}$ and values, $\tau^*_{\rm max}-\tau^*_{\rm residual}$ may be <0. In such cases, the $\tau^*_{\rm max}$ and $\tau^*_{\rm residual}$ combination is thrown out and new values are randomly chosen from the distribution. At $N_{\rm max} = 20$ per cent, unreasonable failure criteria are never noted. At $N_{\rm max} = 50$ per cent, the failure criteria were resampled $\sim 2--4$ times per 100 earthquakes. Other input noise distributions could have been chosen (e.g. lognormal, Weibull). Lacking information to the contrary, a symmetric input noise distribution was desired. Given that realistic noise levels are probably ≤ 20 per cent, a normal distribution is the simplest possible choice.